



The Influence of Transformational Education Prediction on Softskills of Madrasa Student using Data Mining

Andin Aprila Sari
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
andinaprila19@gmail.com

Asma Zulfiah Malik
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
asmazulfiahm@gmail.com

Laela Shintia Alviani
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
laelachintia2006@gmail.com

Annisa Sugi Pramesty
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
annisazyn@gmail.com

Dini Nurul Haq Al-Hidayah
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
dininurul942@gmail.com

Rifka Alia Assyifa
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
rifkaalia.ra@gmail.com

Abstract— Currently, transformative education is oriented towards students' independence in solving a problem they face. In other words, transformative education has its own influence on the soft skills of students. This research was conducted with the aim of providing predictions regarding the effect of transformative education on the soft skills possessed by students. The subjects in this study were students of *Madrasah Aliyah Negeri (MAN) 2 Bandung*, ranging from grade 10 to grade 12. The research method used was the Naïve Bayes and K-Nearest Neighbor algorithm by taking datasets from an independent survey that had been carried out previously. The Naïve Bayes and K-Nearest Neighbor algorithms themselves are included in supervised learning and can be used to predict with a high degree of accuracy. Supervised learning is one of the existing methods in machine learning by means of labeling the data trained by the machine. Testing the data using the Naïve Bayes and K-Nearest Neighbor algorithm obtains predictions that transformative

education affects students' soft skills and produces a very high level of accuracy for the Naïve Bayes algorithm, namely 98% of the 100 existing datasets and accuracy of the K-Nearest Neighbor is 76,67%.

Keywords- *data mining, education, k-nearest neighbor, madrasa, naïve bayes, soft skills*

I. INTRODUCTION

Transformative education strategy is a new perspective on education that was pioneered as an alternative solution to educational problems that have not been fully resolved [1], [2]. Therefore, educational reform is to solve existing problems in the field of education, adjust to the direction of development in the field of education, and bring more hope for future progress [3], [4]. So, education must be dynamic and become a torch in the race and confrontation for social change. The need for transformative education is based on

existing education, we still need various references and guidelines so that education in Indonesia can function properly. Do not be influenced by the surroundings.

Education is often referred to as an effort to improve academic abilities and some abilities that are measurable and can be learned, in this case hard skills. However, through this transformative education, solutions to these problems began to emerge. This is reflected in the orientation of students' transformative education independently solving problems encountered in class and in the environment, students' habits of studying in groups, habits of independence, initiative, creativity and productivity, and planning their lives in the future. In other words, this transformative education has its own tasks and impacts on personality and character, namely soft skills [1].

Education is often referred to as an effort to improve academic abilities and some abilities that are measurable and can be learned, in this case hard skills. However, through this transformative education, solutions to these problems began to emerge. This is reflected in the orientation of students' transformative education independently solving problems encountered in class and in the environment, students' habits of studying in groups, habits of independence, initiative, creativity and productivity, and planning their lives in the future. In other words, this transformative education has its own tasks and impacts on personality and character, namely soft skills.

In addition, madrasas as formal education have a distinction compared to other formal education by prioritizing character education through good religious education (*aqidah, akhlaq, fiqh*) [5]–[9]. The influence of transformative education on the soft skills of madrasah students is an interesting thing to be analysed. Data mining technology can be a solution to find out the effect of transformative education on the soft skills of madrasah students. Therefore, this study aims to determine the effect of transformative education on students' soft skills using a supervised learning approach with the Naive Bayes and K-Nearest Neighbour algorithms.

II. RELATED WORKS

Several studies used Naive Bayes and K-Nearest-Neighbor (KNN) in data mining because those algorithms have good performance. Several related studies are: (1) cervical cancer prediction using Naive Bayes and KNN [10], [11]; (2) sentiment analysis research using Naive Bayes [12]–[15]; (3) recovery prediction of COVID-19 patient using Naive Bayes [16]; (4) Naive Bayes and KNN to predict the issues of divorce case [17], [18]; (5) expert system for many health and disease cases using KNN [19]–[23]; (6) online news portal classification using KNN [24]; and so on.

III. RESEARCH METHODS

A. Research Activities

This research begin with problem identification, data collecting, data mining implementation with Naive Bayes and KNN algorithm, then model evaluation using confusion

matrix. In the problem identification, this research determines the problem formulation of the research. In this case observing the problems related to the prediction of the effect of transformative education on the soft skills in madrasah. The locus of this research in *Madrasah Aliyah Negeri (MAN) 2 Bandung City* students. Existing problems are then analysed to find out how to solve these problems and determine the scope of the problems to be studied. Learn the basic theory from various literature regarding the application of the Naive Bayes and K-Nearest Neighbour methods, through journals to gain basic knowledge for conducting further research.

The systematic procedure used to collect data is by asking questions to respondents. Data collection was carried out using the media google form which was distributed to the research sample, namely MAN 2 Bandung City students. The data collected is student identity data and a scale of suitability of student behaviour with the questions that have been presented. The results of the questionnaire obtained will be used as material for analysing data on the Naive Bayes and K-Nearest Neighbour methods. After the data has been collected, data analysis is carried out to adjust the data process to be processed using the Naive Bayes and K-Nearest Neighbour methods.

Next is the implementation process using Microsoft Excel and Google Colab using Python programming language in the process of building the Naive Bayes and KNN models. This study applies a data mining process which consists of [25], [26]:

1. Data cleaning, is a cleaning or process of removing noise and inconsistent data.
2. Data integration, is a process in which data from various data sources are combined. This process is carried out when more than one data source is used.
3. Data selection, is a data selection process in which the data to be used in the data mining process is taken and leaves unused data.
4. Data transformation, is the process of converting data into a form that can be used in calculating an algorithm.
5. Mining is the process of finding patterns from datasets that are used as a knowledge base, in this case using the Naive Bayes and K-Nearest Neighbor algorithms.
6. Pattern evaluation, is the process of analyzing the results of the mining process using a unit of measure.
7. Knowledge presentation, is a process to display the results of the mining process.

B. Naive Bayes Algorithm

Naive Bayes is an algorithm used for classification that uses Bayes' theorem and assumes that the values between variables are independent (independent) in an output value [27]–[30]. In other words, the presence or absence of a particular variable will not affect or be related to the presence or absence of other variables. This method which is also known as Naive Bayes Classifier implements a supervised object classification technique in the future by assigning class labels to instances/records using conditional probabilities. The term supervised refers to the classification of training data that has been labelled with a class. Equation (1) which

describes Bayes' theorem. In simpler terms, Bayes' Theorem is a simple mathematical formula for finding probabilities when we know certain other probabilities.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Where,

- $P(A|B)$ is Posterior probability of A in condition B (posterior probability)
- $P(B|A)$ = Posterior probability of B in condition A (likelihood)
- $P(A)$ = Prior probability of A (class prior probability).
- $P(B)$ = prior probability of B (predictor prior probability)

C. K-Nearest Neighbor Algorithm

The K-Nearest Neighbor (K-NN) algorithm is a lazy learning technique. K-NN is done by looking for groups of objects in the training data that are closest (similar) to objects in the new data or testing data [31]–[33]. The K-NN algorithm is a method that uses a supervised algorithm. In general, how the KNN algorithm works is as follows [34]:

1. Determine the number of neighbours (K) that will be used for class determination considerations.
2. Calculate the distance from the new data to each data point in the dataset.
3. Take a number of K data with the shortest distance, then determine the class of the new data.

To calculate the distance between two points in the KNN algorithm, the Euclidean Distance method is used which can be used in 1-dimensional space, 2-dimensional space, or multi-dimensional space. 1-dimensional space means that the distance calculation uses only one independent variable, 2-dimensional-space means that there are two independent variables, and multi-dimensional space means that there are more than two variables. In general, the Euclidean distance formula in 1-dimensional space is found in equation (2). Formula (2) can be used if the number of independent variables is only one variable. If there is more than one, it becomes formula (3).

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad (2)$$

$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots} \quad (3)$$

D. Confusion Matrix

Confusion Matrix represents predictions and actual (actual) conditions from data generated by Machine Learning algorithms. Based on the Confusion Matrix, Accuracy, Precision, Recall and Specificity can be determined [35]. Figure 1 describes the parameters that must be considered in the Confusion Matrix.

		Predicted Class	
		Class=Yes	Class=No
Actual Class	Class=Yes	True Positive	False Negative
	Class=No	False Positive	True Negative

Fig 1. Confusion Matrix

Where, True Positives (TP) if the predicted positive value is true, which means that the true class value is "YES" and the predicted class value is also "YES". True Negatives (TN) if the predicted negative value is correctly, it means that the actual class value is "NO" and the predicted class value is also "NO". False Positives (FP) if the actual class is "NO" and the predicted class is "YES". Then, False Negatives (FN) if the actual class is "YES" but the predicted class is "NO".

After understanding these four parameters, Accuracy, Precision, Recall and F1-Score can be calculated. a. Accuracy is the ratio of True predictions (positive and negative) to all data. The formula calculation is available in formula (4). Precision is the ratio of the correct positive predictions compared to the overall positive predicted results. The formula calculation is provided in formula (5). Recall is the ratio of correctly positive predictions compared to all of the correctly positive data. The formula for Recall is available in formula (6). Then, F1 Score is a weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. The formula (7) shows the calculation of F1 Score.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (4)$$

$$Precision = TP / (TP + FP) \quad (5)$$

$$Recall = TP / (TP + FN) \quad (6)$$

$$F1\ Score = (2 \times (Recall \times Precision)) / (Recall + Precision) \quad (7)$$

IV. RESULT AND DISCUSSION

A. Data Preperation Result

The initial step in the data mining process is to clean the data, in other words to ensure that the data obtained does not contain empty values or values that are inconsistent with the values they should be. The cleaning process is done manually on the 100 data you have. Of the 100 data owned, no data has null values or inconsistent values.

Data selection is the process of selecting data based on research variables. The data that will be used in this study are data Name, Gender, Age, Class, T1 (Teamwork Question 1), T2, T3, T4, T5, PS1 (Problem solving Question 1), PS2, PS3, PS4, PS5, C1 (Communication skills questions 1), C2, C3, C4, C5, CT1 (Critical thinking questions 1), CT2, CT3, CT4, and CT5. In the data owned, there is no data that is not included in the research variables.

Data transformation is a process where data is converted into a form that can be processed by an algorithm. There are 3 steps that will be carried out in the data transformation

process to classify using the Naïve Bayes algorithm, including the following:

1. Initializing data that is not yet in numerical form. In this process, Name data as identity is initialized to numbers, from numbers 1-100. In addition, the Gender data is also initialized, with the provision that the "Female" gender is given the initial 1 and the "Male" gender is given the initial 2.
2. Calculates the average value of all scale results on data T1, T2, T3, T4, T5, PS1, PS2, PS3, PS4, PS5, C1, C2, C3, C4, C5, CT1, CT2, CT3, CT4, and CT5. This data is called "Average Scoring" data.
3. Performing discretization, or changing data in continuous form into discrete form. Data with an Average Scoring > 3.5, then the new data which is named "Output" is given the initial 1, which means "has influence". Whereas data with Average Scoring ≤ 3.5 will be given the initial 2, which means "has no influence".

B. Implementation of Naïve Bayes Algorithm

The next stage is the implementation stage of the method described earlier. First, the manual data cleaning process will be carried out, then selecting data, and converting data both into numeric form and into discrete form, so that it can be implemented into the computational process. This process can also be called the training process, where the Naïve Bayes model will be created using 100 datasets which will then be directly tested using the Confusion Matrix to obtain its performance. The result of this stage is a Naïve Bayes model that can already be used to classify and predict results, according to the purpose of this study.

```
[ ] # Menentukan Hasil Prediksi
if y_pred == 1:
    hasil = "Berpengaruh"
elif y_pred == 2:
    hasil = "Tidak Berpengaruh"
else :
    hasil = "Error"

print("Hasil prediksi = ", hasil)

Hasil prediksi = Berpengaruh
```

Fig 1. Result of Naïve Bayes Algorithm

Based on Figure 1, the model with the Naïve Bayes algorithm which is fully listed on Google Colab provides predictive results that transformative education has an influence on students' soft skills, in this case teamwork soft skills, problem solving, communication skills, and critical thinking. Then the next test is carried out. Based on the test results of the model that has been created using the Confusion Matrix, the values shown in Figure 2 are obtained. It can be seen from Figure 2, the accuracy value obtained is 0.98, with a precision value for macro avg of 0.96 and a weighted avg of 0.98, a recall value for macro avg of 0.99 and a weighted avg of 0.98, and an f1-score value for macro avg of 0.97 and weighted avg of 0.98. This shows that the level of accuracy,

precision, recall, and f1-score is high so that the system can be considered good.

```
[ ] # Mencetak Laporan Hasil Klasifikasi
from sklearn.metrics import classification_report
print(classification_report(y, y_predict))
```

	precision	recall	f1-score	support
1	1.00	0.97	0.99	74
2	0.93	1.00	0.96	26
accuracy			0.98	100
macro avg	0.96	0.99	0.97	100
weighted avg	0.98	0.98	0.98	100

Fig 2. Performance of Naïve Bayes Algorithm

C. Implementation of K-Nearest Neighbor Algorithm

This model also has the same data preparation as the model above, but the difference can be seen from the algorithm used. A total of 100 datasets will be used and tested using the Confusion Matrix. Of the 100 datasets, the data is divided into two types, namely 70% for Training Data and 30% for Test Data. The result of this stage is also a KNN model that can be used and a prediction result, according to the purpose of this research.

```
# Perbandingan y_true dan pred
cols = {
    'Y True': y_test,
    'Predict': pred
}
print("Prediksi yang dihasilkan ")
pd.DataFrame(cols)
```

Y True	Predict
0	1
1	1
2	2
3	1
4	2
5	2
6	1
7	2
8	1
9	1
10	1
11	2
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	2
22	2
23	1
24	1
25	1
26	1
27	1
28	1
29	2

Fig 3. Result of KNN Algorithm

```
print("Performance KNN")
accuracy = results_data.test_accuracy.mean()*100
precision = results_data.test_precision.mean()*100
recall = results_data.test_recall.mean()*100

print("Accuracy : %.2f" % accuracy, "%")
print("Precision : %.2f" % precision, "%")
print("Recall : %.2f" % recall, "%")
```

Performance KNN
Accuracy : 76.67 %
Precision : 78.33 %
Recall : 96.67 %

Fig 4. Performance of KNN Algorithm

Based on Figure 3, in contrast to the prediction results given by the Naive Bayes algorithm model in Figure 5, this model provides predictive results according to the row for all 30 Data Tests. In accordance with the data preparation that has been done, the initial 1 is the initial to give information "has influence", while the initial 2 means "has no influence".

Then the performance test was then carried out using the Confusion Matrix, with the test results given in Figure 4.

Based on Figure 4, the accuracy obtained is 76.67% with a precision value of 78.33% and a recall value of 96.67%. Based on the Confusion Matrix parameters, the model with the K-NN algorithm produces the following information: a. There are as many as 76.67% of students from all students who are correctly predicted to be affected by transformative education on their soft skills and are not affected. b. There were as many as 78.33% of students who were correctly predicted to be affected by transformative education on their soft skills of all students who were predicted to be affected. c. There are as many as 96.67% of students predicted to be affected by transformative education on the soft skills they have of all students who are correctly predicted to be affected.

D. Comparison Result between Naïve Bayes and K-Nearest Neighbor Algorithm

Based on the data that has been tested, the calculation results for accuracy, precision, and recall of each algorithm are obtained. The results of testing each algorithm can be seen in Table 1. It can be seen that the performance of the Naive Bayes method is better than the K-Nearest Neighbour (KNN) method. However, this is an invalid result because the amount of data tested for each method is not the same. Thus, research was conducted on other studies to compare the performance of these two methods with the same data. Based on this research, it was concluded that by using 30 test data for both methods, the resulting accuracy for the Naive Bayes algorithm is quite good, this is due to the advantages of the Naive Bayes method itself which is able to classify even though it has little training data for parameter estimation. Whereas the K-Nearest Neighbour method produces low accuracy, this is because the method is not effective if the training data is small [36].

TABLE I. COMPARISON BETWEEN NAÏVE BAYES AND KNN

Method	Accuracy	Precision	Recall
Naïve Bayes	98%	96%	99%
KNN	76.67%	78.33%	96.67%

V. CONCLUSION

In this study, a data mining method was applied in the form of a classification to predict the effect of transformative education on students' soft skills. There are two methods used, namely Naive Bayes and K-Nearest Neighbor. As for the data used in testing the classification method comes from student data MAN 2 Bandung City. Based on the trial it can be concluded that the system is able to predict the effect of transformative education on soft skills with fairly high accuracy. The trials conducted showed that the Naive Bayes method resulted in a higher curation compared to the K-Nearest Neighbor, which was 98%. For further research, you can develop datasets and use other machine learning algorithms.

REFERENCES

- [1] M. Murwanto, "Pendidikan Transformatif dalam Pembentukan Karakter," in *SEMINAR NASIONAL PPG UST*, 2022.
- [2] D. Dawiyatun, "Pendidikan Transformatif," *Islamuna: Jurnal Studi Islam*, vol. 4, no. 2, pp. 290–303, 2017, doi: <https://doi.org/10.19105/islamuna.v4i2.1592>.
- [3] N. Qosim and S. Rohman, "Pendidikan Transformatif: Elektabilitas Santri (kaum sarungan) menjadi Pemimpin Idaman," *BAHTSUNA*, vol. 3, no. 2, pp. 330–346, 2021, doi: <https://doi.org/10.55210/bahtsuna.v3i2.51>.
- [4] D. P. Putri and F. Fitmawati, "Pelaksanaan Pendidikan Transformatif dalam Menghasilkan Pendidik Matematika yang Islami, Profesional, dan Kompetitif," *ARITHMETIC: Academic Journal of Math*, vol. 3, no. 2, pp. 127–146, 2022, doi: <http://dx.doi.org/10.29240/ja.v3i2.3874>.
- [5] Direktorat KSKK Madrasah, "Keputusan Menteri Agama Nomor 183 Tahun 2019 Tentang Kurikulum PAI dan Bahasa Arab pada Madrasah," Jakarta, 2019.
- [6] S. N. Umami, D. S. Oktavia, and A. Sani, "Kekhasan Pendidikan Madrasah Dalam Tinjauan Filosofis," *Al-Astar*, vol. 1, no. 2, pp. 75–87, 2022.
- [7] N. Khamalah, "Penguatan pendidikan karakter di madrasah," *Jurnal kependidikan*, vol. 5, no. 2, pp. 200–215, 2017, doi: <https://doi.org/10.24090/jk.v5i2.2109>.
- [8] A. Nurriqi, "Karakteristik Pendidikan Agama Islam Di Madrasah Prespektif Kebijakan Pendidikan," *BINTANG*, vol. 3, no. 1, pp. 124–141, 2021, doi: <https://doi.org/10.36088/bintang.v3i1.1152>.
- [9] T. Ningsih, "Peran Pendidikan Islam dalam Membentuk Karakter Siswa di Era Revolusi Industri 4.0 di Madrasah Tsanawiyah Negeri 1 Banyumas," *INSANIA: Jurnal Pemikiran Alternatif Kependidikan*, vol. 24, no. 2, pp. 220–231, 2019.
- [10] P. Subarkah, "Perbandingan Kinerja Algoritma CART dan Naive Bayesian untuk Mendiagnosis Penyakit Diabetes Mellitus," 2016.
- [11] E. S. Salim, H. Bindan, E. Pranoto, and A. Dharma, "Analisa Metode Random Forest Tree dan K-Nearest Neighbor dalam Mendeteksi Kanker Serviks," *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, vol. 3, no. 2, pp. 97–101, 2020.
- [12] M. Abbas, K. A. Memon, A. A. Jamali, S. Memon, and A. Ahmed, "Multinomial Naive Bayes classification model for sentiment analysis," *JCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 3, pp. 62–67, 2019.
- [13] R. Rahman, M. A. Masud, R. J. Mimi, and M. N. S. Dina, "Sentiment analysis on bengali movie reviews using multinomial Naive Bayes," in *2021 24th International Conference on Computer and Information Technology (ICCIT)*, IEEE, 2021, pp. 1–6, doi: <https://doi.org/10.1109/ICCIT54785.2021.9689787>.
- [14] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naive Bayes classifier," in *Journal of Physics: Conference Series*, IOP Publishing, 2019, p. 012024, doi: [10.1088/1742-6596/1192/1/012024](https://doi.org/10.1088/1742-6596/1192/1/012024).
- [15] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 2, p. 113, 2018, doi: [10.26418/jp.v4i2.27526](https://doi.org/10.26418/jp.v4i2.27526).
- [16] O. P. Barus and A. Tehja, "Prediksi Kesembuhan Pasien COVID-19 di Indonesia Melalui Terapi Menggunakan Metode Naive Bayes," *Journal Information System Development (ISD)*, vol. 6, no. 2, pp. 59–66, 2021.
- [17] M. Irfan, W. Uriawan, O. T. Kurahman, M. A. Ramdhani, and I. A. Dahlia, "Comparison of Naive Bayes and K-Nearest Neighbor methods to predict divorce issues," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2018, p. 12047.
- [18] I. A. Dahlia, "Perbandingan Metode Naive Bayes dan K-Nearest Neighbor untuk Prediksi Perceraian: Studi kasus Pengadilan Agama Cimahi," Bachelor Thesis, UIN Sunan Gunung Djati Bandung, 2018.
- [19] E. N. Shofia, R. R. M. Putri, and A. Arwan, "Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid Menggunakan Metode K-Nearest Neighbor and Certainty Factor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, e-ISSN*, 2017.
- [20] I. Indrayanti, D. Sugianti, and A. Al Karomi, "Optimasi Parameter K Pada Algoritma K-Nearest Neighbour Untuk Klasifikasi Penyakit Diabetes Mellitus," *Prosiding SNATIF*, pp. 823–829, 2017.
- [21] Y. A. Gerhana, W. B. Zulfikar, A. H. Ramdani, and M. A. Ramdhani, "Implementation of Nearest Neighbor using HSV to Identify Skin

- Disease,” *IOP Conf Ser Mater Sci Eng*, vol. 288, no. 1, p. 012153, 2018, doi: 10.1088/1757-899X/288/1/012153.
- [22] C. S. Fatoni and F. D. Noviandha, “Case Based Reasoning Diagnosis Penyakit Difteri dengan Algoritma K-Nearest Neighbor,” *Creative Information Technology Journal*, vol. 4, no. 3, pp. 220–232, 2018.
- [23] S. Salamun, “Penerapan Algoritma Nearest Neighbor dan CBR pada Expert System Penyimpangan Perilaku Seksual,” *Jurnal Online Informatika*, vol. 2, no. 2, pp. 63–70, 2018.
- [24] S. Asiyah and K. Fithriasari, “Klasifikasi Berita Online Menggunakan Metode Support Vector Machine Dan K-Nearest Neighbor,” *Jurnal Sains dan Seni ITS*, vol. 5, no. 2, 2016, doi: 10.12962/j23373520.v5i2.16643.
- [25] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012. doi: 10.1016/B978-0-12-381479-1.00001-0.
- [26] W. Van Der Aalst, *Process mining: data science in action*, vol. 2. Springer, 2016. doi: <https://doi.org/10.1007/978-3-662-49851-4>.
- [27] D. Buzic and J. Dobsa, “Lyrics classification using Naive Bayes - IEEE Conference Publication,” in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, p. 5. doi: 10.3390/genes6020372.
- [28] I. Destuardi and S. Sumpeno, “Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode Naive Bayes,” *Seminar Nasional Pascasarjana Institut Teknologi Sepuluh Nopember*, 2009.
- [29] C. Slamet, R. Andrian, D. S. Maylawati, Suhendar, W. Darmalaksana, and M. A. Ramdhani, “Web Scraping and Naive Bayes Classification for Job Search Engine,” in *IOP Conference Series: Materials Science and Engineering*, 2018. doi: 10.1088/1757-899X/288/1/012038.
- [30] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa,” *Creative Information Technology Journal*, vol. 6, no. 1, pp. 1–11, 2020, doi: <https://doi.org/10.24076/citec.2019v6i1.178>.
- [31] W. B. Zulfikar and N. Lukman, “Perbandingan Naive Bayes Classifier dengan Nearest Neighbor untuk Identifikasi Penyakit Mata,” *JOIN (Jurnal Online Informatika)*, vol. 1, no. 2, pp. 82–86, 2016.
- [32] M. E. I. Lestari, “Penerapan algoritma klasifikasi Nearest Neighbor (K-NN) untuk mendeteksi penyakit jantung,” *Faktor Exacta*, vol. 7, no. 4, pp. 366–371, 2015.
- [33] D. Syahid, J. Jumadi, and D. Nursantika, “Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (KNN) Berdasarkan Nilai Hue, Saturation, Value (HSV),” *Jurnal Online Informatika*, vol. 1, no. 1, p. 20, 2016, doi: 10.15575/join.v1i1.6.
- [34] M. Mustakim and G. Oktaviani, “Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa,” *SITEKIN: Jurnal Sains, Teknologi dan Industri*, vol. 13, no. 2, pp. 195–202, 2016.
- [35] K. M. Ting, “Confusion Matrix,” in *Encyclopedia of Machine Learning and Data Mining*, 2017. doi: 10.1007/978-1-4899-7687-1_50.
- [36] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan kinerja metode naive bayes dan k-nearest neighbor untuk klasifikasi artikel berbahasa indonesia,” *J. Teknol. Inf. dan Ilmu Komput*, vol. 5, no. 4, 2018, doi: <https://doi.org/10.25126/jtiik.201854773>.